SHORT-PAPER

SMPV: Social Media Prediction for Videos

BO WU, MIT-IBM Watson AI Lab, Cambridge, MA, United States

PEIYE LIU, ByteDance Ltd., Beijing, China

QIUSHI HUANG, University of Surrey, Guildford, Surrey, U.K.

ZHAOYANG ZENG

JIA WANG, Guangdong Pharmaceutical University, Guangzhou, Guangdong, China

BEI LIU, Microsoft Research, Redmond, WA, United States

View all

Published: 27 October 2025

MM '25: The 33rd ACM International Conference on Multimedia
October 27 - 31, 2025
Dublin, Ireland

Conference Sponsors:
SIGMM

https://doi.org/10.1145/3746027.3763757
ISBN: 9798400720352

SMPV: Social Media Prediction for Videos

Bo Wu*
bobbywu.cs@gmail.com
MIT-IBM Watson AI Lab
Cambridge, United States

Peiye Liu*
howard.liu@bytedance.com
ByteDance
San Jose, United States

Qiushi Huang
qiushi.huang@surrey.ac.uk
University of Surrey
Guildford, England

Zhaoyang Zeng
zengzhaoyang@idea.edu.cn
International Digital Economy Academy
Guangzhou, China

Jia Wang
vicky.ee08@nycu.edu.tw
Guangdong Pharmaceutical University
Guangzhou, China

Bei Liu
bei.liu@microsoft.com
Microsoft Research Asia
Beijing, China

Jiebo Luo
jluo@cs.rochester.edu
University of Rochester
Monroe County, United States

Wen-Huang Cheng
wenhuang@csie.ntu.edu.tw
National Taiwan University
Taipei, Taiwan

Abstract

With the explosive growth of video-centric social media platforms, understanding and predicting video popularity has become a crucial problem in both academia and industry. This year, the Social Media Prediction (SMP) Challenge expands its scope by introducing a dedicated video track, shifting the focus from static images to dynamic, multimodal video content. We introduce the Social Media Prediction for Videos (SMPV) task and release a large-scale multimodal benchmark dataset, SMPD-Video, with more than 6K short-form videos, including vision language metadata, user profiles, and popularity labels. This challenge invites global researchers to develop predictive algorithms that integrate spatial-temporal dynamics, multimodal learning, and user-video interactions to forecast video popularity in real-world social temporal streams. With the participation and contribution of top teams around the world, the challenge has seen continuous performance improvements in recent years, driven by technological advancements. SMP Challenge Homepage: www.smp-challenge.com.

CCS Concepts

• **Information systems** → **Web searching and information discovery**; **Multimedia information systems**; • **Computing methodologies** → **Computer vision tasks**; **Natural language processing**.

Keywords

Social Multimedia, Short-form Video Popularity Prediction, Multimodal Learning

MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3763757

ACM Reference Format:
Bo Wu*, Peiye Liu*, Qiushi Huang, Zhaoyang Zeng, Jia Wang, Bei Liu, Jiebo Luo, and Wen-Huang Cheng. 2025. SMPV: Social Media Prediction for Videos. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3746027.3763757

1 Introduction

Social networks have become a global infrastructure for communication, marketing and trend discovery with 5.17 billion users who on average use 6.7 platforms monthly [1, 21]. Advertising spend is projected to reach $219.8 billion in 2024 [24], underscoring the importance of social media prediction [30] for applications such as advertising, recommendation, ranking, and demand forecasting [7, 8, 16, 19, 20, 33]. Within this ecosystem, short-form video has become the dominant medium on platforms such as TikTok, Kwai, and Instagram Reels. With billions of daily views, accurate video popularity forecasting is increasingly critical and enables personalized recommendation, creator monetization, content moderation, and trend detection.

The Social Media Prediction (SMP) Challenge [25, 28, 29] is an annual competition advancing forecasting tasks in social media, bringing together communities in multimedia, AI, social computing, computer vision, and NLP. With the explosive rise of short-form video on platforms such as TikTok and Instagram Reels, and its growing impact on engagement and monetization, this edition introduces a dedicated video prediction track for the first time. This marks a shift from previous years focused on image- and text-based prediction, benchmarking scalable, generalizable algorithms for real-world video popularity forecasting. The key highlights are as follows:

- We introduce the Social Media Prediction for Videos (SMPV) task, which forecasts the future popularity of short-form videos by leveraging multimodal signals, including visual content, temporal context, user profiles, and vision–language metadata.
- We present the SMPV Dataset, a multimodal benchmark comprising 6K video posts from 4.5K users, enriched with tags,

user profiles, thumbnails, captions, hashtags, and detailed interaction statistics.
- The challenge has drawn participation from leading academic and industrial teams worldwide. This year's focus on video prediction reflects the urgent need for scalable and generalizable algorithms to address the rapid growth of video content on social media.

2 Problem Definition

The rapid rise of short-form video platforms such as TikTok and Instagram Shorts has amplified the challenge of information overload, making it increasingly difficult for users to identify relevant content, track emerging trends, or discover influential creators. In this landscape, video-based word-of-mouth signals—captured through likes, shares, comments, and views—play a pivotal role in surfacing popular and engaging content.

To tackle this challenge, we introduce the task Social Media Popularity Prediction for Videos (SMPP-Video), which seeks to forecast the future popularity of video posts by modeling multimodal signals and contextual information [26, 27, 30]. The task is to predict the **popularity score** s of a **short-form video post** v published by a user u at time t. Consider a set of n triplet data user-post-time in social media streams, each user-post pair (u, v) associated with a specific sharing time t. The sequence of user-post can be represented as

$$S = \{(u_1, v_1)_{t1}, \ldots, (u_n, v_n)_{tn}\}, \quad \text{where} \quad t_1 \leq \cdots \leq t_n.$$

The popularity score represents the level of interactions the post receives on a social media platform, s quantifies user engagement, such as views, likes, shares, or comments; in this challenge, we adopt Tiktok interactions as the primary measures of popularity since it directly reflects audience attention. Because social engagement is highly skewed, the **log-normalized popularity distribution** [27] spans from a few views to millions, underscoring the need for robust predictive modeling.

This task requires teams to design innovative algorithms that can automatically predict how popular video posts will become, using multimodal content, posting time, and user-context signals in the fast-changing landscape of social media [14, 32].

3 Dataset Overview

We develop the SMPD-Video(Social Media Prediction Dataset - Video)[1], a large-scale, multimodal benchmark centered on short-form video content, designed to reflect the diversity and complexity of real-world social media platforms. The dataset contains approximately 4.5K users and 6K video-centric posts, spanning over 24 months of temporal engagement data. Each entry is enriched with user profiles, vision-language metadata, 120 semantic category tags, and fine-grained timestamped popularity indicators, such as view counts. To preserve the temporal dynamics inherent in real-world scenarios, we divide the data chronologically into training and testing sets based on posting time. Participating teams are challenged to develop innovative prediction algorithms that can accurately model content popularity in a streaming video environment, with evaluations based on both accuracy and correlation.

[1] http://smp-challenge.com/dataset

Table 1: SMPD-Video: Summary of the Dataset.

Metrics	Statistics	
	Train	Test
Number of Posts	4.05×10^3	2.01×10^3
Temporal Coverage of Posts	24 months	
Number of custom tags	4×10^4	
Number of users	4.5×10^3	
Number of 3^{rd} level categories	120	

4 Evaluation

For objective evaluation, all methods are tested on a held-out portion of the SMPD dataset, kept unpublished during the competition to ensure fairness. We use *Mean Absolute Percentage Error (MAPE)* and *Spearman Rank Correlation (SRC)* as primary metrics, providing complementary measures of predictive accuracy and ranking consistency. The final leaderboard is derived through a composite scheme: teams are first ranked by MAPE, then the overall ranking is determined by balancing both metrics to ensure fair comparison across precision and robustness of predictions.

5 Research Progress

In recent years, research on social media prediction has advanced rapidly, driven by innovations in embedding learning [6, 15, 18, 23], semantic modeling [9, 12], user–post interaction learning [2, 10], and multimodal visual–textual representations. These directions have begun to uncover the inherent complexity of social media data by modeling high-level correlations. Techniques such as attention mechanisms, recurrent architectures, and sliding-window models [3–5, 12, 31] have yielded steady performance gains.

The competition of this year attracted a diverse set of academic and industrial teams, who advanced the field beyond traditional paradigms [11, 13, 17, 22]. Many of the best solutions embraced vision-language contrastive learning, cross-modal semantic alignment, dual-stream transformers, and high-order adaptation, effectively capturing the rich multimodal dynamics of video content. These methods leverage intricate relationships among visual appearance, textual metadata, and user-interaction signals over time. The winning team set a new benchmark by achieving the best SRC and MAE scores, surpassing the first-place results in previous years. In particular, the top SRC improved by more than 10%, highlighting both the competitiveness of the current track and the growing maturity of multimodal video forecasting approaches. Full rankings and detailed results are reported on the official leaderboard.

6 Conclusion

We introduced the Social Media Prediction for Videos (SMPV) task and dataset, advancing multimodal approaches to short-form video popularity forecasting. This year's results show that vision–language alignment, dual-stream models, and temporal modeling yield substantial gains. SMPV establishes a foundation for future work on long-term temporal dynamics, cross-platform transfer, and interpretable multimodal forecasting. We hope this task fosters deeper collaboration across multimedia, natual language, and social computing communities, driving progress in personalized recommendation, trend detection, and creator engagement.

References

[1] Linda C. Ashar, JD. 2024. Social Media Impact: How Social Media Sites Affect Society. https://www.apu.apus.edu/area-of-study/business-and-management/resources/how-social-media-sites-affect-society/

[2] Junhong Chen, Dayong Liang, Zhanmo Zhu, Xiaojing Zhou, Zihan Ye, and Xiuyun Mo. 2019. Social media popularity prediction based on visual-textual features with xgboost. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 2692–2696.

[3] Weilong Chen, Feng Hong, Chenghao Huang, Shaoliang Zhang, Rui Wang, Ruobing Xie, Feng Xia, Leyu Lin, Yanru Zhang, and Yan Wang. 2020. Curriculum Learning for Wide Multimedia-Based Transformer with Graph Target Detection. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 4575–4579.

[4] Weilong Chen, Chenghao Huang, Weimin Yuan, Xiaolu Chen, Wenhao Hu, Xinran Zhang, and Yanru Zhang. 2022. Title-and-Tag Contrastive Vision-and-Language Transformer for Social Media Popularity Prediction. In *Proceedings of the 30th ACM International Conference on Multimedia*. 7008–7012.

[5] Xiaolu Chen, Weilong Chen, Chenghao Huang, Zhongjian Zhang, Lixin Duan, and Yanru Zhang. 2023. Double-Fine-Tuning Multi-Objective Vision-and-Language Transformer for Social Media Popularity Prediction. In *Proceedings of the 31st ACM International Conference on Multimedia*.

[6] Keyan Ding, Ronggang Wang, and Shiqi Wang. 2019. Social media popularity prediction: A multiple feature fusion approach with deep neural networks. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 2682–2686.

[7] Emilio Ferrara, Roberto Interdonato, and Andrea Tagarelli. 2014. Online Popularity and Topical Interests Through the Lens of Instagram. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (Santiago, Chile) *(HT '14)*. 24–34.

[8] Xiangnan He, Ming Gao, Min-Yen Kan, Yiqun Liu, and Kazunari Sugiyama. 2014. Predicting the popularity of web 2.0 items based on user comments. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*. 233–242.

[9] Chih-Chung Hsu, Li-Wei Kang, Chia-Yen Lee, Jun-Yi Lee, Zhong-Xuan Zhang, and Shao-Min Wu. 2019. Popularity prediction of social media based on multi-modal feature mining. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 2687–2691.

[10] Chih-Chung Hsu, Chia-Ming Lee, Xiu-Yu Hou, and Chi-Han Tsai. 2023. Gradient Boost Tree Network based on Extensive Feature Analysis for Popularity Prediction of Social Posts. In *Proceedings of the 31st ACM International Conference on Multimedia*.

[11] Chih-Chung Hsu, Chia-Ming Lee, Yu-Fan Lin, Yi-Shiuan Chou, Chih-Yu Jian, and Chi-Han Tsai. 2024. Revisiting Vision-Language Features Adaptation and Inconsistency for Social Media Popularity Prediction. In *Proceedings of the 32st ACM International Conference on Multimedia*.

[12] Chih-Chung Hsu, Wen-Hai Tseng, Hao-Ting Yang, Chia-Hsiang Lin, and Chi-Hung Kao. 2020. Rethinking relation between model stacking and recurrent neural networks for social media prediction. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 4585–4589.

[13] Wenhao Hu, Weilong Chen, Weimin Yuan, Yan Wang, Shimin Cai, and Yanru Zhang. 2024. Dual-Stream Pre-Training Transformer to Enhance Multimodal Learning for Social Media Prediction. In *Proceedings of the 32st ACM International Conference on Multimedia*.

[14] Shoubin Kong, Qiaozhu Mei, Ling Feng, Fei Ye, and Zhe Zhao. 2014. Predicting bursts and popularity of hashtags in real-time. In *Proceedings of the 37th international ACM SIGIR Conference on Research & development in Information Retrieval*. 927–930.

[15] Xin Lai, Yihong Zhang, and Wei Zhang. 2020. HyFea: Winning Solution to Social Media Popularity Prediction for Multimedia Grand Challenge 2020. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 4565–4569.

[16] Cheng Li, Yue Lu, Qiaozhu Mei, Dong Wang, and Sandeep Pandey. 2015. Click-through Prediction for Advertising in Twitter Timeline. In *Proceedings of KDD*.

[17] Yu-Shi Lin and Anthony J.T. Lee. 2024. MMF: Winning Solution to Social Media Popularity Prediction Challenge 2024. In *Proceedings of the 32st ACM International Conference on Multimedia*.

[18] Shijian Mao, Wudong Xi, Lei Yu, Gaotian Lü, Xingxing Xing, Xingchen Zhou, and Wei Wan. 2023. Enhanced CatBoost with Stacking Features for Social Media Prediction. In *Proceedings of the 31st ACM International Conference on Multimedia*.

[19] Travis Martin, Jake M Hofman, Amit Sharma, Ashton Anderson, and Duncan J Watts. 2016. Exploring limits to prediction in complex social systems. In *Proceedings of International World Wide Web Conference (WWW)*.

[20] Gabor Szabo and Bernardo A Huberman. 2010. Predicting the popularity of online content. *Commun. ACM* 53, 8 (2010), 80–88.

[21] WebFX Team. 2024. 100+ Social Media Statistics You Should Know for 2024. https://www.webfx.com/social-media/statistics/

[22] Mingsheng Tu, Tianjiao Wan, Qisheng Xu, Xinhao Jiang, Kele Xu, and Cheng Yang. 2024. Higher-Order Vision-Language Alignment for Social Media Prediction. In *Proceedings of the 32st ACM International Conference on Multimedia*.

[23] Kai Wang, Penghui Wang, Xin Chen, Qiushi Huang, Zhendong Mao, and Yongdong Zhang. 2020. A Feature Generalization Framework for Social Media Popularity Prediction. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 4570–4574.

[24] Chloe West and Chloe West. 2024. 50+ Must-know Social Media Marketing Statistics for 2024. https://sproutsocial.com/insights/social-media-statistics/

[25] Bo Wu, Wen-Huang Cheng, Peiye Liu, Bei Liu, Zhaoyang Zeng, and Jiebo Luo. 2019. SMP Challenge: An Overview of Social Media Prediction Challenge 2019. In *Proceedings of the 27th ACM International Conference on Multimedia*.

[26] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, and Tao Mei. 2016. Time Matters: Multi-scale Temporalization of Social Media Popularity. In *Proceedings of the 2016 ACM on Multimedia Conference (ACM MM)* (Amsterdam, The Netherlands).

[27] Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Huang Qiushi, Li Jintao, and Tao Mei. 2017. Sequential Prediction of Social Media Popularity with Deep Temporal Context Networks. In *International Joint Conference on Artificial Intelligence (IJCAI)* (Melbourne, Australia).

[28] Bo Wu, Peiye Liu, Wen-Huang Cheng, Bei Liu, Zhaoyang Zeng, Jia Wang, Qiushi Huang, and Jiebo Luo. 2023. SMP Challenge: An Overview and Analysis of Social Media Prediction Challenge. In *Proceedings of the 31th ACM International Conference on Multimedia*.

[29] Bo Wu, Peiye Liu, Qiushi Huang, Zhaoyang Zeng, Jia Wang, Bei Liu, Jiebo Luo, and Wen-Huang Cheng. 2024. SMP Challenge Summary: Social Media Prediction Challenge. In *Proceedings of the 32nd ACM International Conference on Multimedia* (Melbourne VIC, Australia) *(MM '24)*. Association for Computing Machinery, New York, NY, USA, 11442–11444.

[30] Bo Wu, Tao Mei, Wen-Huang Cheng, and Yongdong Zhang. 2016. Unfolding Temporal Dynamics: Predicting Social Media Popularity Using Multi-scale Temporal Decomposition. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI)* (Phoenix, Arizona).

[31] Kele Xu, Zhimin Lin, Jianqiao Zhao, Peicang Shi, Wei Deng, and Huaimin Wang. 2020. Multimodal deep learning for social media popularity prediction with attention mechanism. In *Proceedings of ACM international conference on Multimedia (ACM MM)*. 4580–4584.

[32] Jaewon Yang and Jure Leskovec. 2011. Patterns of Temporal Variation in Online Media. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)*.

[33] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. 2015. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1513–1522.